

# Homework 3: Gephi

## Interactive visualization and exploration for all kinds of networks

*Due 11:59pm EEST December 13, 2019*

### General Instructions

Your answers should be as concise as possible.

**Submission instructions:** You should submit your answers and code in a compressed directory uploaded via <https://eclass.aueb.gr>

*Submitting answers:* Prepare a short description of your answers and personal choices on this homework in a single PDF file named *hw3.pdf*

*Submitting code:* Prepare a compressed directory containing a *.gephi* file with your network and the directory created using the *SigmaExporter* plugin.

### Problem

Gephi is an open-source visualization and exploration software for all kinds of graphs and networks.

#### 1 Visualize the communities of commonly occurring adjectives and nouns

This homework requires that you import a network of common adjective and noun adjacencies for the novel "David Copperfield" by Charles Dickens into Gephi [1]. Nodes represent the most commonly occurring adjectives and nouns in the book. Edges connect any pair of words that occur in adjacent position in the text of the book.

After you import the network, you should resize and color its nodes by measuring statistics and applying filters to highlight certain attributes, visualize

it using a drawing algorithm, and export it to a web application that enables interactive network display.

To begin with, you must install a plugin that is available in Gephi named *SigmaExporter*. To do so, click on Tools → Plugins and select and install the aforementioned plugin from the *Available Plugins* tab.

Then you have to import the network specified in the files ‘nodes.csv’ and ‘edges.csv’ that is provided to you with this homework.

Next you ought to perform the following:

- Measure the degrees of all nodes and resize the nodes according to their degree.
- Find:
  - the network’s diameter,
  - the individual with the largest betweenness centrality,
  - the communities existing in the network.
- Color the nodes according to the communities you discovered and **comment on** the observations you make after examining these communities. Provide some examples of nouns that ended up in the same community although they represent very different things? What do you think is the reason they are in the same community?
- Select a word in the network apply different colors to the edges of:
  - this node,
  - the nodes in its ego network with depth 1,
  - the nodes in its ego network with depth 2, and
  - the remaining nodes.
- Apply a layout than enhances the visualization of your network (ForceAtlas, Fruchterman Reingold, Yifan Hu and their variations are excellent choices).

In this step you have to make sure that the resulting visualization is aesthetically pleasing with regard to the following:

- Nodes are not overlapping,
- Sizes of nodes clearly indicate differences in in-degree of nodes and are appropriate with regard to the edge thickness and the distance of nodes, and
- Colors are highly contrasted.

Finally, you need to export the network using the *SigmaExporter* plugin. If the latter is installed properly, you can use it through File → Export → Sigma.js template. Remember to fill in the title and author fields appropriately.

**HINTS:**

- Coloring the nodes of the various ego networks can be achieved using the ego network filter, starting from the whole network and progressing until depth 1.
- You can use the Data Laboratory to import the network.

**CAREFUL:** Your report should clearly specify the word you chose, along with its degree and PageRank value.

## References

- [1] M. E. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.